Geometrical aspects of chiral anomalies in the overlap.

Herbert Neuberger

Department of Physics and Astronomy
Rutgers University
Piscataway, NJ 08855–0849

Abstract

The set of one dimensional lowest energy eigenspaces used to construct the overlap induces a two form on gauge orbit space which is the locally exact curl of Berry's connection. If anomalies do not cancel, examples of two dimensional closed sub-manifolds of orbit space are produced over which the integral of the above two form does not vanish. Based on these observations, a natural definition of covariant currents is obtained, a simple way to calculate chiral anomalies on the lattice is found, and indications for how to construct an ideal regularization of chiral gauge theories are seen to emerge.

1. Introduction.

In the continuum, on a compact Euclidean space-time manifold, chiral anomalies can be understood and evaluated from solely geometric considerations [1]. On a finite lattice it would appear that these insights have to be lost, being somehow restored in the continuum limit. In this paper the overlap will be shown to provide a geometrical interpretation for chiral anomalies directly on lattices approximating a continuum torus. With this understanding it will become clearer why theories where anomalies cancel between different multiplets are fundamentally different in the overlap approach from those where the anomalies do not cancel. This insight holds directly on the lattice and does not appeal to continuum physics or perturbative concepts.

The essence of the overlap [2,3,4,5] is the association of the continuum chiral determinant, viewed as a line bundle over the space of gauge field configurations factored by gauge transformations, with a line bundle of ground states of a certain bilinear fermionic Hamiltonian over the space of all compact link variables factored by lattice gauge transformations. (More precisely, the lattice complex line bundle consists of a collection of projections of the ground states along a fixed vector, denoted by $|v_+ >$ in the appendix.) This association is imperfect because the space of gauge orbits in the continuum is disconnected while the analogous space on the lattice is connected. This imperfection is reflected by the need to excise from the space of lattice gauge orbits those for which the dimensionality of the ground state eigenspace exceeds unity because of accidental degeneracies. The set of excised configurations has zero weight in the measure induced by the Haar volume element per link and per site for link parallel transporters and for gauge transformations, respectively. The excised configurations are "exceptional" in the sense adopted in early studies of gauge field topology on the lattice [6]. The removal of the exceptional configurations leaves behind a space of lattice gauge orbits that no longer is connected, but the number of components if finite.

The chiral determinant vanishes over all connected components of the space of lattice orbits except one. The Hamiltonians commute with fermion number and the fermion number of $|v_+ >$ is definite. Thus, the projection along $|v_+ >$ vanishes whenever the ground state has a fermion number different from that of $|v_+ >$. In any construction of chiral gauge theories one first focuses on the one component where the chiral determinant does not vanish. This component is a connected, continuous space of gauge orbits and this work is restricted to it. To define the chiral determinant for these backgrounds a smooth section of the line bundle of projected vacuum eigenspaces would be needed.

We shall find two kinds of obstacles to the construction of a gauge invariant chiral

determinant: the first consists of obstructions which need to mutually cancel and the second is a residue remaining after the cancelation. Both obstacles can be understood in the framework of Berry's phase [7]. The basic object will be a two form over the space of gauge potentials, which is locally exact and given by the curl of Δj, the difference between the covariant and the consistent currents [8]. This abelian curvature plays a central role also in other regularizations of the chiral gauge theories [10], but in the overlap it acquires a simple geometrical interpretation, becoming Berry's curvature for a certain line bundle.

As usual, instead of working over the space of orbits we start by working over the space of gauge connections, or link variables. Obstructions tell us that factoring by gauge transformations would not produce single valued functions over the space of orbits. In section 2 we show that the usual definitions of the consistent and covariant currents [8], j^{cons}, j^{cov} and Δj are consistent with j^{cons} being the variation of the lattice overlap w.r.t the gauge field and Δj being Berry's connection associated with a certain set of finite vectors $|v>$ parameterized by the gauge fields [9]. We then proceed to show that Berry's curvature $d\Delta j$ is a non-perturbatively, gauge invariantly regularized version of the antisymmetric two form $Z(\delta_1 A, \delta_2 A, A)$ of [10]. In sections 3 and 5 we show in dimensions 2 and 4 that the known continuum expressions for $Z(\delta_1 A, \delta_2 A, A)$ can be determined directly from Berry "monopole" singularities. In section 4 it is shown that in a certain two dimensional case, for which anomalies cancel, the remaining lattice artifact terms in $Z(\delta_1 A, \delta_2 A, A)$ can be diminished by smooth deformations of the constructions of the $|v>$ line bundle. If anomalies do not cancel no such deformations can affect a certain component of the total curvature; this is the component that survives in the continuum but it separates cleanly from the other contributions, already on the lattice. In section 6 the usual Brillouin-Wigner phase choice is shown to also possess a certain geometrical meaning. It is suggested that this insight explains why the BW phase convention turned out to obey several desirable symmetries [3]. Several conclusions and conjectures are described in section 7. Appendix A contains a definition of the lattice overlap in any even dimension. Appendix B clarifies the compact notation of section 2 in the lattice context.

Let me briefly discuss the generality of the results: Sections 3 through 5 deal with abelian gauge backgrounds. The results of sections 3 and 5 extend to the non-abelian case by embedding. Section 4 presents an explicit abelian example of an improved choice of Hamiltonians. It demonstrates that by identifying the "monopole" sources of anomalies in statistically important backgrounds, one can reduce lattice artifacts in anomaly free cases. While the example is abelian and two dimensional, the principle behind it is general and applies to any even dimension and any compact gauge group. The principle is explained in section 2.

2. Δj and Berry's phase in the overlap.

This section has two parts. In the first I review the concepts of covariant and consistent currents in the continuum and introduce a compact notation to emphasize the essentials of the structures. In the second part I describe lattice objects corresponding to the continuum currents of the first part.

2a. Continuum.

Since our independent variables are the continuum gauge fields I find it easier to start by choosing a notation that makes this obvious: Replace the fields $A^a_\mu(x)$ by real coordinates ξ_α. The index α takes all the values taken by the triplet (μ, a, x). An infinitesimal gauge transformation is parameterized by a function of the pair (a, x), and will be denoted by ω_i, where i replaces (a, x). The consistent current (to be defined below) is a functional of the gauge field, a function of x and has components labeled by μ and a. It will be denoted by $j^{\text{cons}}_\alpha(\xi)$. Clearly, it can be viewed as a one form $j^{\text{cons}} = j^{\text{cons}}_\alpha d\xi_\alpha$ over the space coordinatized by the ξ_α. Under a finite gauge transformation g, our coordinates ξ_α get replaced by $(\xi^g)_\beta$:

$$(\xi^g)_\beta = h_\beta(g) + \mathcal{D}^{-1}_{\alpha\beta}(g)\xi_\alpha. \tag{2.1}$$

The first term is an inhomogeneous global shift and the second is a homogeneous linear transformation. The equation is nothing but the usual gauge transformation. We shall only need it for small variations $\delta\xi$, where the inhomogeneous term drops out. We have

$$\frac{\partial}{(\xi^g)_\alpha}\xi_\beta = \mathcal{D}_{\alpha\beta}(g), \tag{2.2}$$

where $\mathcal{D}_{\alpha\beta}(g)$ is real.

A gauge invariant function $I(\xi)$ obeys:

$$I(\xi) = I(\xi^g). \tag{2.3}$$

Functions $\psi_\alpha(\xi)$ (a one form $\psi_\alpha(\xi)d\xi_\alpha$) that transform as the gradient $\partial_\alpha I$ (dI), are said to transform covariantly:

$$\psi_\alpha(\xi^g) = \mathcal{D}_{\alpha\beta}(g)\psi_\beta(\xi). \tag{2.4}$$

The matrices $\mathcal{D}_{\alpha\beta}(g)$ represent the action of a gauge transformation g on the $\psi_\alpha(\xi)$ objects in a covariant manner.

Associated with the Hamiltonian $\mathcal{H}^-$ (see appendix) is a Hilbert space of finite dimension, $\mathcal{V}$, providing a Fock representation for a set of canonical fermionic creation and

annihilation operators a, $a^\dagger$. Gauge transformations g are unitarily represented on this Hilbert space by $G(g)$ in a ξ_α independent way.

Let us assume now that we have a regularization that produces a chiral determinant $D(\xi)$. The consistent current is defined by:

$$j_\alpha^{\text{cons}}(\xi) = \partial_\alpha \log D(\xi). \tag{2.5}$$

Consistency [8] simply means,

$$\partial_\beta j_\alpha^{\text{cons}} = \partial_\alpha j_\beta^{\text{cons}}, \tag{2.6}$$

or

$$dj^{\text{cons}} = 0, \tag{2.7}$$

as $j^{\text{cons}} = d \log D(\xi)$.

An infinitesimal gauge transformation acts on functions of ξ_α in a way dictated by the form of $(\xi^g)_\alpha$ (eq. (2.1)),

$$\partial_i = (X_{i\alpha} + Y_{i\alpha\beta}\xi_\beta)\partial_\alpha \tag{2.8}$$

with ξ-independent X and Y. Y comes from the linear part of $\mathcal{D}$ expanded around $g \equiv 1$ and proportional to ω_i. Hence ([8]),

$$\partial_\alpha \partial_i - \partial_i \partial_\alpha = Y_{i\beta\alpha}\partial_\beta, \tag{2.9}$$

which implies:

$$\partial_\alpha(\partial_i \log D) - \partial_i j_\alpha^{\text{cons}} = Y_{i\beta\alpha} j_\beta^{\text{cons}}. \tag{2.10}$$

The infinitesimal form of eq. (2.4) is $\partial_i \psi_\alpha + Y_{i\beta\alpha}\psi_\beta = 0$. If anomalies are absent, j_α^{cons} transforms covariantly because $\partial_i \log D = 0$, the Y implementing the familiar commutator. A non-vanishing anomaly implies a noncovariant transformation law for the consistent current in the non-singlet, nonabelian case. Bardeen and Zumino (BZ) [8] show by explicit construction that there always exists a one form Δj, polynomial (in ξ), local in space-time, such that $j^{\text{cons}} + \Delta j$ transforms covariantly.

In summary, the anomalous, non-abelian situation in the continuum is as follows: There are two currents, j^{cons} and j^{cov}. j^{cov} is not the variation of any function, but is gauge covariant. j^{cons} is the derivative of a function (the regulated chiral determinant) but is not gauge covariant. Given j^{cons} we could reconstruct the regulated determinant, but the lack of gauge covariance of j^{cons} makes the reconstructed function break gauge invariance. When anomalies cancel j^{cons} and j^{cov} are equal. In that case the total reconstructed

5

determinant is gauge invariant. So, when we go to the lattice, we should focus our attention on Δj, the difference between j^{cons} and j^{cov}. We wish to understand on the lattice directly why in the anomalous case it is unavoidable that $\Delta j \neq 0$. Then, we wish to see that when anomalies cancel it no longer is unavoidable that $\Delta j \neq 0$. To make $\Delta j = 0$ one would need to tune the lattice overlap Hamiltonian matrices (for definitions see Appendix A). I present an argument that the obstruction preventing $\Delta j = 0$ *on the lattice* in the anomalous case disappears if the *continuum algebraic conditions* for anomaly cancelation hold. The argument is geometrical, goes to the heart of the matter, and is supported by abelian examples. In the abelian case the compactness of the group is crucial.

2b. Lattice.

On the lattice, in a nonperturbative framework, at a finite cutoff, with compact gauge fields, Δj will not be polynomial and one does not regularize just the *variation* of the chiral determinant, as is sometimes done in the continuum [10]; the determinant itself is regularized. One way or another, when defining the chiral determinant, one always deals with a determinant line bundle over the space of ξ's (before attempting to factor out gauge transformations). The overlap is no exception, only now one has one dimensional spaces that are naturally embedded in one common larger space. This allows us to compare the one dimensional spaces over different points ξ. Therefore, one has a natural split of the variation in a vector at ξ induced by deforming ξ to $\xi + \delta\xi$: one part of the variation is the "real" change in the spaces spanned by each vector and the other is the "irrelevant" change along the fibers.

Since $< v_+ |$ is taken to be ξ independent (see Appendix A) the variation of the regulated chiral determinant is directly related to the variation of a vector in $\mathcal{V}$:

$$\delta < v_+ | v_- > = < v_+ | \delta v_- > . \tag{2.11}$$

Henceforth we shall suppress the subscript minus and shall replace $|v_- >$ by $|v >$. (Similarly, the Hamiltonian superscript will be dropped.) The split we described earlier is:

$$\partial_\alpha | v > = (\partial_\alpha | v >)_\perp + | v > < v | \partial_\alpha | v >, \tag{2.12}$$

where $< v | [(\partial_\alpha | v >)_\perp] = 0$ since $| v >$ is normalized by definition.

This split was noted already in [11], but the roles of the two terms were misidentified, based on an example calculation that turned out to be wrong, producing an incorrect coefficient.[f1] This error precluded further development until, about one year ago, S.

[f1] Equation 3.70 in ref. [11] has the wrong overall sign and the first term in the round brackets should be deleted.

Randjbar-Daemi and J. Strathdee [5], starting from scratch arrived at the same split, but this time correctly identified the roles of the pieces.[f2] The first term in equation (2.12), measuring the distance between the spaces, corresponds to j^{cov}, the covariant form of the current. The second term gives Δj. (Obviously, the sum gives j^{cons}.)

Let me first show explicitly that j^{cov} so identified indeed transforms covariantly. Under $\xi \to \xi^g$, a replacement of the parameters in $\partial_\alpha \mathcal{H}$, we have

$$\mathcal{H}(\xi^g) = G^\dagger(g)\mathcal{H}(\xi)G(g), \tag{2.13}$$

leading to

$$(\partial_\alpha \mathcal{H})(\xi^g) = \mathcal{D}_{\alpha\beta}(g)G^\dagger(g)(\partial_\beta \mathcal{H})(\xi)G(g). \tag{2.14}$$

The symbols ξ, α, ∂_α, are natural generalizations of their continuum counterparts in the previous subsection. Readers who find this confusing are referred to Appendix B. Ordinary perturbation theory tells us that:

$$(\partial_\alpha |v>)_\perp(\xi) = \frac{1}{\mathcal{H}(\xi) - E_0(\xi)} |w(\xi)>, \quad \text{where}$$
$$|w(\xi)> \equiv \big[<v(\xi)|\partial_\alpha \mathcal{H}(\xi)|v(\xi)> - \partial_\alpha \mathcal{H}(\xi)\big]|v(\xi)> . \tag{2.15}$$

The above formula is well defined since $<v(\xi)|w(\xi)> = 0$. Now, eq. (2.13) implies:

$$|v(\xi^g)> = e^{i\phi_v(\xi,g)}G^\dagger(g)|v(\xi)> . \tag{2.16}$$

The phase factor is arbitrary. Also, $G(g)|v_+> = e^{i\alpha(g)}|v_+>$, where the phase provides a one dimensional representation of the group of gauge transformations. A simple calculation,

$$(\partial_\alpha |v>)_\perp(\xi^g) =$$
$$G^\dagger(g)\frac{1}{\mathcal{H}(\xi) - E_0(\xi)}(\xi)G(g)\mathcal{D}_{\alpha\beta}(g)\big[<v(\xi^g)|G^\dagger(g)\partial_\beta \mathcal{H}(\xi)G(g)|v(\xi^g)> -$$
$$G^\dagger(g)\partial_\beta \mathcal{H}(\xi)G(g)\big]|v(\xi^g)>$$
$$= e^{i\phi_v(\xi,g)}\mathcal{D}_{\alpha\beta}(g)G^\dagger(g)[(\partial_\alpha|v>)_\perp](\xi), \tag{2.17}$$

shows covariance:

$$\frac{<v_+|[(\partial_\alpha|v>)_\perp](\xi^g)}{<v_+|v>(\xi^g)} = \mathcal{D}_{\alpha\beta}(g)\frac{<v_+|[(\partial_\beta|v>)_\perp](\xi)}{<v_+|v>(\xi)}. \tag{2.18}$$

[f2] They also computed various quantities, including the afore mentioned coefficient; their equation (24) provides a correct replacement for equation 3.73 in [11].

Note the important disappearance of the unknown phase due to the cancelation between the numerator and denominator. This is Fujikawa's view [12] of the gauge non-invariance being restricted to the "fermionic measure" at work. To compute j^{cov} in the lattice overlap one does not need to make a phase choice. Thus, the covariant currents are defined naturally and gauge covariantly.[f3] The same goes for the covariant anomaly: it has no dependence on the phase choice in the overlap.[f4]

In the continuum we know that for non-singlet non-abelian anomalies Δj vanishes if and only if we have anomaly cancelation, because if there are anomalies the consistent and covariant currents can't be equal. So, anomaly cancelation in this case is equivalent to the vanishing of Δj. In the overlap, the main impediments to arrange for Δj to vanish identically by deforming the overlap lattice Hamiltonians used in the construction of the states $|v_\pm >$ will be seen to disappear if anomalies cancel. The natural and direct definition of Δj and its curl in the overlap easily extends to the abelian case, unlike a definition based on (2.10). Just like the covariant current, the curl of Δj is also defined in a gauge invariant manner.

The formula for Δj in the overlap is simple:

$$\Delta j_\alpha = < v|\partial_\alpha|v > \tag{2.19}$$

Δj depends on the phase choice for $|v>$, but, the curl of Δj_α $(d\Delta j)$ does not; if $d\Delta j \neq 0$ there is no way a phase choice could eliminate Δj. Whether or not there is a nonzero curl depends only on the Hamiltonian. It is useful now to recall Berry's phase [7]. Clearly, Δj is nothing but the Berry connection [9], while the curl is Berry's curvature [7, 9].

The curl of Δj was analyzed in [10] in the continuum. The curl, (an abelian field strength over the space of ξ_α), in components $\mathcal{F}_{\alpha\beta} = \partial_\alpha \Delta j_\beta - \partial_\beta \Delta j_\alpha$, is more conveniently manipulated after contraction with two arbitrary "vectors" δ_α^1 and δ_β^2. The quantity $\mathcal{F}_{\alpha\beta}\delta_\alpha^1\delta_\beta^2$ is denoted in references [10] by a functional $Z(\delta_1 A, \delta_2 A, A)$ where $\delta_1 A_\mu(x)$ and $\delta_2 A_\nu(y)$ play the role of δ_α^1 and δ_β^2. To us the most important aspect of the analysis in [10] is that the curvature is finite and apparently unambiguous in the continuum regularization adopted there which required no gauge breaking even at intermediary steps. Thus, the curvature is potentially just as fundamental as the anomaly itself. On the lattice, the overlap provides a nonperturbative framework to realize the same situation.

[f3] This is particularly useful for QCD applications [13], but a more detailed discussion would take us too far off track here.

[f4] Essentially, this is why in the original proposal of Kaplan [14] it was possible to compute the anomaly without any apparent ambiguity: the outcome turned out to be the covariant anomaly [15].

We know that there are typically two sources to Berry's curvature: One consists of a "smeared" collection and the other of "monopole" singularities [7]. The "monopole" singularities cannot be made to go away by small deformations of the Hamiltonian. But the smeared component of the source can.

3. Two dimensions: Berry's curvature and the abelian anomaly.

In this section we shall see that anomalies indeed correspond to the "monopoles" identified by Berry, and that anomaly cancelation corresponds to them canceling each other. The connection between anomalies and Berry's phase is different from previous relations described in the literature [16] and is specific to the overlap framework, but not to the particular form of regularization within which the overlap is implemented.

We are considering a Weyl fermion in two Euclidean dimensions in an abelian external $U(1)$ gauge field. Space-time is taken to be a flat torus obtained by identifying the opposite boundaries of a square. The torus is replaced by a mesh of small squares and covered by L^2 such plaquettes.

A family of gauge backgrounds is chosen to consist of a collection of constant gauge potentials. Thus, we are concentrating on a flat torus embedded in the space of gauge orbits. There are no exceptional points (in the sense of the introduction) on this torus, so it is a compact, smooth manifold. Our objective is to show that Berry's curvature associated with the one form $< v|dv >$ on this manifold integrates to a non-zero value one could associate with degeneracies "nearby". The degeneracy points lie outside the space of ξ's, and assume the role of "monopole" sources. Since we are dealing with the integral of a locally exact form, its value is quantized, and small deformations of the Hamiltonian cannot remove the singularity. The total strength of the singularities is q^2 where q is the integral $U(1)$-charge of the Weyl fermion, here assumed to be a left mover. For a right mover we get $-q^2$. For the integral of the total curvature over the torus to vanish the well known anomaly cancelation condition $\sum_R q_R^2 = \sum_L q_L^2$ must hold. In that case, the "monopoles" cancel each other out.

We replace the variables $U_\mu(x)$ by e^{ih_μ} for each x in eqs. (A.3). By gauge invariance the variables h_μ are periodic with period $\frac{2\pi}{L}$ each.

The matrix H block diagonalizes to two by two blocks in Fourier space and the ground state of $\mathcal{H}$ is obtained by filling the negative energy state corresponding to each block. We label the momenta and the associated blocks and states by a two component integral vector n, $(n_\mu = 0, ..., L-1, \ \mu = 1, 2)$.

9

For $h_\mu = 0$ we have:

$$p_n = \frac{2\pi}{L}n, \qquad H_n = \begin{pmatrix} \frac{1}{2}\hat{p}_n^2 - m & i\bar{p}_n^1 - \bar{p}_n^2 \\ -i\bar{p}_n^1 - \bar{p}_n^2 & m - \frac{1}{2}\hat{p}_n^2 \end{pmatrix}. \qquad (3.1)$$

We use $\bar{p}_\mu = \sin p_\mu$ and $\hat{p}_\mu = 2\sin\frac{p_\mu}{2}$. For arbitrary h_μ we simply need to replace every p_n by $p_n + h$.

The curvature we wish to compute is given in second quantized language by

$$\mathcal{F}_{\alpha\beta} = <\partial_\alpha v | \partial_\beta v> - <\partial_\beta v | \partial_\alpha v>, \qquad (3.2)$$

where $\alpha = (\mu, x)$ and $\beta = (\nu, y)$. For our background we get

$$\sum_{x,y} \mathcal{F}_{\alpha\beta} \equiv f_{\mu\nu} = <\frac{\partial v}{\partial h_\mu} | \frac{\partial v}{\partial h_\nu}> - <\frac{\partial v}{\partial h_\nu} | \frac{\partial v}{\partial h_\mu}>. \qquad (3.3)$$

The two form f, $f = \frac{1}{2}f_{\mu\nu}dh_\mu dh_\nu$, is taken over the h-torus. Define the numbers $\tilde{f}(m)$ for integral two dimensional vectors m:

$$\tilde{f}(m) = \left(\frac{L}{2\pi}\right)^2 \int_{|h_\mu| \leq \frac{\pi}{L}} e^{-iLh\cdot m} f(h). \qquad (3.4)$$

$\tilde{f}(0)$ will be seen to be quantized and controlled solely by the "monopoles"; absence of anomalies is equivalent to the cancelation of all the $\tilde{f}(0)$ terms among all fermion species. Eq. (3.4) can be inverted:

$$f_{12}(h) = \sum_{m \in Z^2} e^{iLh\cdot m} \tilde{f}_m. \qquad (3.5)$$

It is straightforward to see that the equation $da = f$ for an unknown one-form $a(h)$ has solutions over the torus if and only if $\tilde{f}(0) = 0$. The undefined part of the one form a that can be written as $d\Phi$ can be eliminated by a phase choice for the ground state; the rest can only be eliminated by deforming the matrix H.

Writing out explicitly the Slater determinant wave function for the ground state of $\mathcal{H}$ in terms of single particle wave functions one easily derives

$$f_{\mu\nu}(h) = \sum_n \Big[\frac{\partial u^\dagger(p_n + h)}{\partial h_\mu} \frac{\partial u(p_n + h)}{\partial h_\nu} - (\mu \leftrightarrow \nu) \Big], \qquad (3.6)$$

where $u(p_n)$ is a normalized negative energy eigenstate of H_n. Clearly, such an object carries a phase choice, and we wish to make it explicit that f does not depend on this phase choice. For this we need the projector $P_n(h)$ on the appropriate eigenspace, and an

expression for f in terms of $P_n(h)$'s only. The (easily proven) required expression can be found in [17]:

$$< \delta_1 u | \delta_2 u > -(1 \leftrightarrow 2) = tr(\delta_2 P P \delta_1 P - (1 \leftrightarrow 2)), \tag{3.7}$$

where $P = |u><u|$ and $P^2 = P$. All we need at the moment for proceeding is that the $H_n(h)$ are two by two hermitian traceless matrices, and therefore:

$$P_n(h) = \frac{1}{2}(1 - \vec{w}_n(h) \cdot \vec{\sigma}). \tag{3.8}$$

Here, $\vec{\sigma}$ is the usual triple of Pauli matrices and the real three vectors $\vec{w}_n(h)$ have unit length. Up to a positive prefactor we have $H_n(h) \propto \vec{w}_n \cdot \vec{\sigma}$. Since $tr(\delta_1 P \delta_2 P) = tr(\delta_2 P \delta_1 P)$, we can write:

$$< \delta_1 u | \delta_2 u > -(1 \leftrightarrow 2) = tr(\delta_2 P(1/2 - P)\delta_1 P) - (1 \leftrightarrow 2). \tag{3.9}$$

Simple algebra now produces:

$$< \delta_1 u | \delta_2 u > (p_n + h) - (1 \leftrightarrow 2) = \frac{i}{2} \vec{w}_n \cdot \delta_1 \vec{w}_n \times \delta_2 \vec{w}_n. \tag{3.10}$$

We recognize the appearance of the infinitesimal area element on the surface of the sphere $\vec{w}_n^2 = 1$.

$$f_{12}(h) = \frac{i}{2} \sum_n \vec{w}_n \cdot \frac{\partial \vec{w}_n}{\partial h_1} \times \frac{\partial \vec{w}_n}{\partial h_2}. \tag{3.11}$$

The expression above needs to be integrated over the h-torus. For any function F we have:

$$\int_{|h_\mu| \leq \frac{\pi}{L}} d^2 h \sum_n F(p_n + h) = \int_{|\theta_\mu| \leq \pi} F(\theta). \tag{3.12}$$

Therefore (see(3.4)),

$$\tilde{f}(0) = i \frac{L^2}{2\pi} N, \tag{3.13}$$

where N is the number of times the torus $|\theta_\mu| \leq \pi$ wraps around the sphere $w^2 = 1$ under the map $\theta \longrightarrow \vec{w}$, explicitly given by:

$$\begin{pmatrix} \frac{1}{2}\hat{\theta}^2 - m & i\bar{\theta}^1 - \bar{\theta}^2 \\ -i\bar{\theta}^1 - \bar{\theta}^2 & m - \frac{1}{2}\hat{\theta}^2 \end{pmatrix} \equiv E(\theta)\vec{w}(\theta) \cdot \vec{\sigma}. \tag{3.14}$$

By definition, $E(\theta) \geq 0$ and $\vec{w}^2(\theta) = 1$. It is easy to see that the mapping near the south pole of the $\vec{w}$-sphere is one to one, so $N = 1$.

The source of the winding is easily identified if one considers deforming the parameter m to negative values. In that case the south pole is never reached so $N = 0$. When $m = 0$

11

there is a degeneracy at $\theta_\mu = 0$. This degeneracy is a Berry "monopole". When m goes to $-\infty$ the torus gets mapped into a single point on the sphere. Considering the images of the torus as a function of the parameter m we acquire the picture of a monopole traveling from the "outside" of the torus into its "inside" as m is increased from $-\infty$ through zero, and once it crosses into the "interior" the winding number changes from zero to unity. We conclude that to obtain the quantity $\tilde{f}(0)$ we only need to survey the various zero energy degeneracy points one can induce by varying also the parameter m. These degeneracies occur in the first quantized Hamiltonian matrix and in the ground state of the second quantized Hamiltonian operator simultaneously. Only degeneracies at zero energy of H play the role of "monopoles" for the overlap.

Until now, we dealt with a fermion of unit charge. If the fermion has charge q, N gets replaced by $q^2 N$. Actually, one has q^2 distinct monopoles, not one monopole of strength q^2.

To change handedness we know that we should replace in the overlap formula the lowest energy states by the highest energy ones. This simply amounts to replacing the projectors P by $1 - P$, inducing a sign switch in N, and having the expected effect on the anomaly.

The calculation showed that, regardless of the phase choice, i.e. before any gauge non-invariant step has been taken, we can conclude that the antisymmetric tensor $\mathcal{F}_{\alpha\beta}$ has a constant piece given by:

$$\pm q^2 \frac{i}{2\pi} \epsilon_{\mu\nu} \delta_{xy} \equiv \pm q^2 \frac{i}{2\pi} \mathcal{E}_{\alpha\beta}. \tag{3.15}$$

This equation is consistent with the following continuum expression:

$$Z(\delta_1 A, \delta_2 A, A) = \pm i \frac{q^2}{2\pi} \epsilon_{\mu\nu} \int d^2 x \, \delta_1 A_\mu(x) \delta_2 A_\nu(x). \tag{3.16}$$

$\mathcal{E}_{\alpha\beta}$ is a covariant (trivially, as it is constant), antisymmetric tensor in ξ-space. Such a tensor is available because of the existence of $\epsilon_{\mu\nu}$ in two dimensional space-time. Equation (3.16) agrees with [5,10].

Note that the calculation was done on a finite lattice. The torus that we used was not in momentum space, as the latter is discrete. Also note that the anomaly is traced to a quantized integer already on the lattice, so there is no question about the continuum limit. This is in general line with the basic philosophy of the overlap, namely that anomalies should appear as phenomena completely divorced from ultraviolet effects.

4. Two dimensions: An example of partial improvement.

If anomalies do not cancel there is no way to proceed to eliminate the curl of Δj. If this cannot be done, there is no hope to find a phase choice for the overlap that would render a vanishing Δj, and a gauge invariant chiral determinant. On the other hand, we would like to be able to arrange for Δj to vanish if anomalies do cancel, because then the regulated determinant would be gauge invariant, since its derivatives w.r.t. to the gauge fields would be given by the covariant current while the latter is defined in a gauge invariant manner and is curl free. I don't expect a simple solution for all possible gauge backgrounds. In this section we focus on a subset of backgrounds over which the curl of the Δj corresponding to an abelian anomaly free two dimensional chiral model is set to zero by a simple adjustment in the regularization.

The model is a favorite of overlap work [18]: It contains one periodic left mover of charge 2 and four right movers of charge 1. We know that to consistently quantize the model on a two torus, one needs to pick the four right movers to obey the set of four boundary conditions (PP), (PA), (AP) and (AA). Here, (PA) for example, means periodic in direction 1 and anti-periodic in direction 2.

We still restrict our attention to the h-torus. The total curvature will be the algebraic sum of the curvatures contributed by each one of the five fermions. Per fermion these contributions have the form of equation (3.6). Generically it looks like:

$$f_{\mu\nu}(h) = \epsilon_{\mu\nu} \sum_n \hat{f}(p_n + h). \tag{4.1}$$

Let us assume that the Hamiltonians for each fermion species are regularized on similar square lattices. If the size of the lattice is L_F for fermion F of charge q_F and the argument h in (4.1) is h_F, the combination $\frac{L_F h_F}{q_F}$ must be F-independent. The function $\hat{f}$ for each fermion depends only on the form of the Hamiltonian. Let us take the simple case that all Hamiltonians are picked of the same form. Let the charge 2 fermion live on a lattice of size L, and have charge $q = 2$. Let the charge 1 fermions all live on lattices of the same size L_1. We choose to work with h-variables periodic with period $\frac{2\pi}{L}$. To implement the boundary conditions we introduce variables n_{PP}, n_{PA}, n_{AP}, n_{AA}:

$$n_{PP} \equiv n; \quad n_{PA} \equiv n + (0, 1/2); \quad n_{AP} \equiv n + (1/2, 0); \quad n_{AA} \equiv n + (1/2, 1/2). \tag{4.2}$$

The total curvature is then:

$$f_{\mu\nu}^{\text{total}}(h) = \epsilon_{\mu\nu} \Bigg[\sum_{n;\ n_\mu=0,1,\ldots L-1} \hat{f}(\frac{2\pi}{L}n + 2h) -$$

$$\sum_{n;\ n_\mu=0,1,\ldots L_1-1} \hat{f}(\frac{2\pi}{L_1}n_{PP} + \frac{L}{L_1}h) -$$

$$\sum_{n;\ n_\mu=0,1,\ldots L_1-1} \hat{f}(\frac{2\pi}{L_1}n_{PA} + \frac{L}{L_1}h) - \qquad (4.3)$$

$$\sum_{n;\ n_\mu=0,1,\ldots L_1-1} \hat{f}(\frac{2\pi}{L_1}n_{AP} + \frac{L}{L_1}h) -$$

$$\sum_{n;\ n_\mu=0,1,\ldots L_1-1} \hat{f}(\frac{2\pi}{L_1}n_{AA} + \frac{L}{L_1}h) \Bigg].$$

Clearly, the choice $L_1 = L/2$ produces exact cancelation and zero total curvature on the h-torus. The set of boundary conditions for the charge 1 fermions work precisely so that each one of the four monopoles associated with the charge 2 fermion is individually canceled by a monopole associated with one specific charge 1 fermion. The choice of $L_1 = L/2$ was implemented in previous work [19], but its impact on the phase of the overlap was not mentioned.

It is not necessary to have a strictly gauge invariant chiral determinant for all gauge backgrounds; one can rely on gauge averaging and the Foerster, Nielsen, Ninomiya mechanism instead [20]. Nevertheless, in the numerical simulations carried out for this model [19] it was important to use a definition of the overlap that was close to "ideal" at least for the gauge field configurations that carry significant statistical weight when the lattice coupling becomes large and continuum is approached. Since constant gauge fields are not suppressed in the continuum limit, getting close to a gauge invariant definition on the h-torus discussed above was not only nice but actually necessary in practice.

The reason for the necessity is not fundamental, but has to do with the inability of Monte Carlo techniques to deal reliably with complex measures. The phase of the integrand in the integral estimated by the Monte Carlo procedure was absorbed into the observable. Typically this would lead to disastrous results since wild fluctuations of the phase are expected, so impractical long simulation times would be needed to dig the signal out from under the noise. (Actually, this may not be possible even in principle because of roundoff errors reflecting the finite number of digits used to represent floating point numbers on a computer.) In the particular model of [18] the continuum limit is exactly soluble, and tells us that the chiral determinant is actually positive. Thus all the phases we would see are lattice artifacts and, in principle, would be completely eliminated in an

"ideal" regularization.

In practice, the cancelation of the curvature on the h-torus combined with the BW phase choice (see section 6 for a definition) turn out to be sufficient. As mentioned above, the FNN mechanism indicates that it is not necessary to have an "ideal" regularization; it is the practical aspects of Monte Carlo integration that make it necessary to go some distance towards an ideal regularization. Of course, it would be nice to know that there exists, in principle, a choice of Hamiltonians and other regularization details (something one could call a perfect, or ideal, overlap - see section 12, bottom of page 380, in [3]) that provide strict gauge invariance for anomaly free theories. On the other hand, it is also important to know (at least on the lattice) for sure whether, in principle, any *fine* tuning is needed in order to regularize anomaly free chiral gauge theories. My feeling, based on the FNN mechanism and available numerical evidence to date is that no *fine* tuning is needed in principle. At present, this opinion does not appear to be widely shared among workers in the field.

5. Four dimensions: Berry's curvature and the abelian anomaly.

In four dimensions we again consider a charged left handed Weyl fermion interacting with a $U(1)$ gauge field. First, we need to identify a sub-manifold of ξ-space on which it is easy to show that the curl of Δj cannot be made to vanish by small smooth deformations of the Hamiltonian matrices.

In section 2 we learned that the relevant obstructions can be found by varying m and looking for zero energy degeneracies in the Hamiltonian matrix. Counting the parameter m as one of the dimensions, in addition to ξ, the above degeneracies are co-dimension three points. The easiest case would be a two dimensional sub-manifold in the space of gauge orbits on which the Hamiltonian globally breaks up into two by two blocks. By varying m we can search for "monopoles" associated with any one of the 2×2 blocks. We pick two directions, say 1 and 2, and make the corresponding components of the gauge potentials in those directions constant. This generates a torus over which Berry's curvature could integrate to a non-zero value. For this we need a covariant (constant) antisymmetric tensor with two indices, α and β, as before. On the torus there are no space-time variables we can use to produce antisymmetry so we need to use the $\epsilon_{\mu\nu\rho\sigma}$ tensor. To absorb two of its indices the simplest is to introduce a constant magnetic field in the 3 and 4 directions. Therefore, we pick the 3 and 4 components of the gauge potential independent of x_1 and x_2, but with the dependence on x_3 and x_4 chosen so as to generate a constant (and hence

quantized) magnetic field through all 3-4 plaquettes. These are instanton configurations from the view point of a 3-4 two dimensional world, and they have been extensively studied before [11,21].

In summary, we pick the following family of backgrounds:

$$
\begin{aligned}
U_3(x) &= \begin{cases} 1, & x_3 \neq L - 1 \\ e^{-i\frac{2\pi}{L}x_4}, & x_3 = L - 1 \end{cases} \\
U_4(x) &= e^{i\frac{2\pi}{L^2}x_3} \\
U_1(x) &= e^{ih_1} \\
U_2(x) &= e^{ih_2}
\end{aligned}
\tag{5.1}
$$

We assume for the time being unit fermion charge and that the system is defined on an L^4 torus. The backgrounds are parameterized by a two dimensional torus consisting of points labeled by h.

We proceed to describe a basis in which the Hamiltonian matrix corresponding to these backgrounds becomes 2×2 block-diagonal. Since there is no dependence on x_1 and x_2 we first go to Fourier space in these variables. We shall denote the momenta by p_n just as in the previous sections. These momenta are two dimensional.

Employing notations given in the appendix (using the lower left corner of the list in (A.4)), H is explicitly given by:

$$
H = (B - m)\sigma_3 \otimes \mathbf{1} + iW_3\sigma_2 \otimes \sigma_3 - iW_4\sigma_1 \otimes \mathbf{1} + i\sigma_2 \otimes [W_1\sigma_1 + W_2\sigma_2].
\tag{5.2}
$$

One should view H as a square matrix acting on a space of dimension $4L^4$ where the L^4 factor comes from the L^2 labels (x_3, x_4) and the L^2 labels p_n. The factor 4 comes from the two two dimensional spaces made explicit in (5.2). H is block diagonal in the n indices and its $4L^2 \times 4L^2$ diagonal blocks will be labeled by H^n. Also, $B_1 + B_2 = \frac{1}{2}\hat{p}^2$.

Associated with each n-block introduce two new $2L^2 \times 2L^2$ matrices, $H_{\pm}^n$:

$$
H_{\pm}^n = \pm iW_3\sigma_2 - iW_4\sigma_1 + (B_3 + B_4 - m + \frac{1}{2}\hat{p}^2)\sigma_3.
\tag{5.3}
$$

The dependence on p_n and h comes in through the quantity $\hat{p}^2$, where p stands for the combination $p_n + h = \frac{2\pi}{L}n + h; \ 0 \leq h_\mu < \frac{2\pi}{L}$. Any p with $0 \leq p_\mu < 2\pi$ uniquely identifies an h and an n. As before, the dependence is really only on the combination $p_n + h$ and this will allow us to use equation (3.12).

The matrices $H_{\pm}^n$ are two hermitian $d = 2$ Hamiltonians in $d = 2$ instanton backgrounds. Let us introduce their $2L^2$-dimensional eigenvectors:

$$
H_{\pm}^n \psi_{n\pm}^A = E_{n\pm}^A \psi_{n\pm}^A.
\tag{5.4}
$$

16

Since $H^n_+ = -\sigma_2 H^n_- \sigma_2$ we can choose $\psi^A_{n+} \equiv \psi^A_n$ and $\psi^A_{n-} \equiv \hat{\psi}^A_n = \sigma_2 \psi^A_n$, with $E^A_{n+} = E^A_n$ and $E^A_{n-} = -E^A_n$.

Introduce now the following $4L^2$-dimensional vectors:

$$\phi^A_n = \psi^A_n \otimes \uparrow, \qquad \hat{\phi}^A_n = \hat{\psi}^A_n \otimes \downarrow . \tag{5.5}$$

The index A takes $2L^2$ values, the index n takes L^2 values and the set $\{\phi^A_n, \hat{\phi}^A_n\}$ constitutes an orthonormal basis of the $4L^4$ dimensional space H is acting on. The two dimensional spinors $\uparrow$ and $\downarrow$ are the ± 1 eigenvectors of the σ_3 factor in $\mathbf{1} \otimes \sigma_3$. In view of W_1 and W_2 being proportional to the unit matrix,

$$
\begin{aligned}
H^n \phi^A_n &= E^A_n \phi^A_n + (iW^n_1 + W^n_2)\hat{\phi}^A_n \\
H^n \hat{\phi}^A_n &= - E^A_n \hat{\phi}^A_n + (iW^n_1 - W^n_2)\phi^A_n .
\end{aligned}
\tag{5.6}
$$

Therefore, H^n has nonzero matrix elements only between states that carry the same index A. For each A, H^n reduces to a two by two matrix H^n_A given by:

$$H^n_A = \begin{pmatrix} E^A_n & iW^n_1 + W^n_2 \\ iW^n_1 - W^n_2 & -E^A_n \end{pmatrix} \tag{5.7}$$

Since $\det H^n_A = -(E^A_n)^2 - |W^n_1|^2 - |W^n_2|^2$ for any A and any n, we know that there are no exceptional configurations on our torus, so it is a compact and smooth manifold just as it was in the section 2.

The dependence on h and n enters as follows:

$$iW^n_1 \pm W^n_2 = -\bar{p}_1 \pm i\bar{p}_2, \tag{5.8}$$

and

$$E^A_n = g_A(m - \frac{1}{2}\hat{p}^2). \tag{5.9}$$

The functions $g_A(\mu)$ are simply the eigenvalues of the $d = 2$ Hamiltonian $iW_3\sigma_2 - iW_4\sigma_1 + (B_3 + B_4 - \mu)\sigma_3$ in an instanton background, viewed as function of a mass parameter μ and do not explicitly depend on n or h. We know [3,21] that as long as $\mu < 0$, $g_A(\mu)$ is bounded away from zero. Thus, for $m < 0$ all blocks H^n_A stay non-degenerate. However, from numerical work [3,21], we know that, for exactly one A, $g_A(\mu)$ crosses zero once as μ goes from negative to a positive value less than 2. To get a degeneracy in (5.7) we also need, in addition to $E^A_n = 0$, $\bar{p} = 0$ there. Among the four values satisfying $\bar{p} = 0$ only $p = 0$ also sets $\hat{p}^2 = 0$. For this value, we see now that a variation of m will take us through a degeneracy and an irremovable contribution to Δj. For the other values, $\frac{1}{2}\hat{p}^2 \geq 2$, and

the argument of g_A stays negative as long as $m \leq 2$; thus potential "monopoles" and "anti-monopoles" associated with fermion doublers are avoided.

If the fermion has integral charge q there will be q values of the index A for which $g_A(\mu)$ would cross zero. Each such crossing is in the same direction and the combined effect (at infinite L one expects the crossings to occur simultaneously in m) can be viewed as that of one monopole of strength q. The analysis of the previous sections shows that the two dimensional system will then see q^2 such monopoles, so the total contribution goes as q^3. In four dimensions we can decide that all Weyl fermions are taken as left handed, and then, to cancel anomalies we need $\sum_F q_F^3 = 0$, as expected. A switch in handedness is equivalent to a switch in the sign of q.

It remains to identify the normalization. If we increased the strength of the constant magnetic field through the 3-4 planes by an integral factor l the number of crossings and, consequentially, Berry's curvature would increase l-fold. Thus, the curvature is proportional to the total flux through a 3-4 plane divided by 2π, its smallest quantum. Putting this together we arrive, in continuum notation, at

$$
\begin{aligned}
Z(\delta_1 A, \delta_2 A, A) = &\pm i \frac{q^3}{4\pi^2} \epsilon_{\mu\nu\rho\sigma} \int d^4 x \, \delta_1 A_\mu(x) \delta_2 A_\nu(x) \partial_\rho A_\sigma(x) = \\
&\pm i \frac{q^3}{8\pi^2} \epsilon_{\mu\nu\rho\sigma} \int d^4 x \, \delta_1 A_\mu(x) \delta_2 A_\nu(x) F_{\rho\sigma}.
\end{aligned}
\tag{5.10}
$$

This agrees with [5,10]. It is also convenient to write the answer in terms of ordinary forms,

$$
\delta_1 A(x) = \delta_1 A_\mu(x) dx_\mu, \quad \delta_2 A(x) = \delta_2 A_\mu(x) dx_\mu, \quad F(x) = \frac{1}{2} F_{\mu\nu} dx_\mu dx_\nu,
$$
$$
\epsilon_{\mu\nu\rho\sigma} dx_1 dx_2 dx_3 dx_4 = dx_\mu dx_\nu dx_\rho dx_\sigma,
\tag{5.11}
$$

as:
$$
Z(\delta_1 A, \delta_2 A, A) = \pm i \frac{q^3}{(2\pi)^2} \int \delta_1 A(x) \delta_2 A(x) F(x).
\tag{5.12}
$$

The generalization to d dimensions [10] for this abelian case is

$$
Z(\delta_1 A, \delta_2 A, A) = \pm i \frac{q^2}{2\pi(\frac{d}{2} - 1)!} \int \delta_1 A(x) \delta_2 A(x) \left(\frac{qF}{2\pi} \right)^{\frac{d}{2} - 1} (x).
\tag{5.13}
$$

The above equation indicates the following interpretation: The torus we are working on is spanned by constant gauge fields in directions 1 and 2. The factor $\frac{q^2}{2\pi}$ comes from q^2 monopoles. The rest of the gauge fields are picked to create constant minimal size magnetic fields through each one of the planes $(3, 4)$, $(5, 6)$, etc. This explains the appearance of

the field strength to power $\frac{d}{2} - 1$. Each field strength comes multiplied by the charge q and divided by 2π because this is the basic quantum of flux. There are $\frac{d}{2} - 1$ factors in all possible orders, but there should be only one, and this is fixed by the prefactor $\frac{1}{(\frac{d}{2}-1)!}$. Using equation (A.4) and the above analysis in four dimensions it should be easy to check how this works out explicitly in $d > 4$.

The setup of fields we employed is very similar to the one used in continuous Minkowski space by Nielsen and Ninomiya [22] to provide a simple interpretation for anomalies. The main difference is in the first two dimensions: in Minkowski space one uses a real constant electric field while here, in Euclidean space, we used only constant gauge potentials. This difference stems from anomalies in Euclidean space having to do exclusively with phases.

It would be instructive to see how much of the present analysis can be reproduced in the nonabelian case, where one can employ constant gauge potentials in all directions. This might also be useful to large N reduced models.

6. The Brillouin-Wigner phase choice.

The Brillouin-Wigner (BW) phase choice is defined for gauge configurations with a non-degenerate $|v>$, where, in addition, $<v_0|v>\neq 0$. $|v_0>$ is a carefully chosen state, typically the ground state of $\mathcal{H}$ for the gauge background $U_\mu(x) = 1$ [3,19]. The phase of the overlap is fixed by requiring $<v_0|v>$ to be positive.

Actually, an equally appropriate description of the phase choice would be as Pancharatnam's phase choice [23]. In the context of the Poincare representation of light polarizations, two beams $|A>$ and $|B>$ were defined by Pancharatnam to be in phase if $<A|B>>0$. He observed that if $|B>$ is in phase with $|A>$ and $|C>$ with $|B>$ then $|C>$ need not be in phase with $|A>$. This clearly reflects an underlying non vanishing curvature. It implies that we cannot arrange for all $|v>$'s to be in phase with each other.

Berry [23] showed that if $|A>$ was parallel transported to $|B>$ using Berry's connection on the Poincare sphere, along the shortest geodesic, $|A>$ and $|B>$ ended up in phase. If we found a specific regularization that made $d\Delta j = 0$ for a specific anomaly free theory we could arrange for all $|v>$'s to be in phase and the BW phase condition would also have been satisfied.

Note the related fact that the BW phase condition can be defined from an extremum principle, namely, requiring to maximize $\||v_0> + |v>\|^2$. In the generic case the extremum principle would guarantee compatibility of the BW phase choice with symmetries in the sense analyzed in detail in [3] on a case by case basis.

7. Discussion.

Sections 2 and 4 present probably the simplest "derivation" of anomalies (including their normalization) in any lattice regularization. The geometric insight also produces a clean definition of covariant currents associated with global or local symmetries acting on the fermions. The separation of sources for Berry's curvature into "monopoles" and the rest may provide the first step in establishing in principle the existence of an "ideal" lattice overlap regularization for any anomaly free gauge theory, chiral or not.

A potential example of an "ideal" lattice regularization might be obtained by replacing $B + \gamma_\mu W_\mu$ in (A.2) by the recently discussed fermionic actions in [24] and use the new Hamiltonians in (A.5) to define the $|v>$-line bundle. One would also need to exactly map the one dimensional line in lattice action space followed by the marginally relevant nonlinear flow associated with a fermion mass and use the theory with the negative mass sign (assuming exact parity invariance for the pure gauge part of the action) and replace m accordingly. At least naively, the above suggestion brings the lattice overlap "closest" (in the sense of violation of conformal invariance) to a continuum overlap, and, if the latter were really defined outside perturbation theory, it should work as an ideal regularization as long as the mass parameter is perceived as infinite on physical scales. Of course, more work is needed to test this suggestion.

Unfortunately, the definition of the Wilson-Dirac operator we would use, as given in [24], is implicit, requiring the solution of a complicated non-linear RG recursion relation on an infinite lattice, since arbitrarily separated fermions eventually become coupled (albeit weakly). Only after the solution is obtained can we go to a finite lattice by factoring translations appropriately. It hasn't been proven yet that a unique solution to the recursion relation exists, and even if it does, it remains unclear whether the solution provides a matrix whose entries are smoothly dependent on the arbitrary link variables. For the considerations of this paper, a smooth dependence of the Hamiltonian matrices on the group valued gauge fields is absolutely essential.

It might be worthwhile to mention here that our four dimensional backgrounds, for any value of the mass parameter in the ranges we considered, can easily be deformed to (abelian) configurations for which the associated Wilson-Dirac operators have a pair of exact zero modes. The deformation is obtained by introducing a parameter κ in the exponents of $U_{3,4}(x)$ in equation (5.1) and reducing κ from 1 towards zero. It is guaranteed that this kind of special configurations will be encountered for some $0 < \kappa < 1$, depending on the mass m. These configurations have what would amount to a singularity in the continuum, and there is no reason to associate the fermion zero modes with four dimensional

instantons. This may be of interest to present day numerical QCD work.

I feel that the geometric insight presented in this paper reveals one of the deeper aspects of the overlap formalism.

Acknowledgments.

Appendix A.

Based on [3], we know that the overlap for a chiral fermions in even d dimensions, interacting with gauge field in a compact group G_s, is defined as follows: On the lattice, the chiral determinant is replaced at the regulated level by the overlap of two fermionic many-body states. These are the ground states of two bilinear Hamiltonians,

$$\mathcal{H}^{\pm} = a^{\dagger} H^{\pm} a, \qquad (A.1)$$

with all indices suppressed. The matrices $H^{\pm}$ are obtained from

$$H(m) = \gamma_{d+1} \left[B - m + \gamma_{\mu} W_{\mu} \right] \qquad (A.2)$$

with $H^{+} = H(-\infty)$, $H^{-} = H(m_0)$ and $0 < m_0 < 2$. The infinite argument for H^{+} can be replaced by any finite positive number, but the equations are somewhat simpler with the above choice [18,19,25]. The matrices W_{μ} and $B = \sum_{\mu} B_{\mu}$ are given below:

$$
\begin{aligned}
(W_{\mu})_{xi,yj} &= \frac{1}{2} [\delta_{y,x+\hat{\mu}} U_{\mu}(x) - \delta_{x,y+\hat{\mu}} U_{\mu}^{\dagger}(y)]^{ij}, \\
(B_{\mu})_{xi,yj} &= \frac{1}{2} [2\delta_{xy}\mathbf{1} - \delta_{y,x+\hat{\mu}} U_{\mu}(x) - \delta_{x,y+\hat{\mu}} U_{\mu}^{\dagger}(y)]^{ij}.
\end{aligned}
\qquad (A.3)
$$

x, y ($x_{\mu} = 0, 1, .., L - 1$) are sites on the lattice and i, j are color indices. The matrices $U_{\mu}(x)$ are G_s link variables on the lattice, associated with direction μ and site x. The γ_{μ} are Euclidean Dirac matrices in d dimensions. We choose the following chiral basis, with

alternating real symmetric and imaginary antisymmetric, matrices:

$$
\begin{aligned}
\gamma_1 &= \sigma_1 \otimes \sigma_1 \otimes \sigma_1 .... \otimes \sigma_1 \otimes \sigma_1 \otimes \sigma_1 \\
\gamma_2 &= \sigma_1 \otimes \sigma_1 \otimes \sigma_1 .... \otimes \sigma_1 \otimes \sigma_1 \otimes \sigma_2 \\
\gamma_3 &= \sigma_1 \otimes \sigma_1 \otimes \sigma_1 .... \otimes \sigma_1 \otimes \sigma_1 \otimes \sigma_3 \\
\gamma_4 &= \sigma_1 \otimes \sigma_1 \otimes \sigma_1 .... \otimes \sigma_1 \otimes \sigma_2 \otimes \mathbf{1} \\
\gamma_5 &= \sigma_1 \otimes \sigma_1 \otimes \sigma_1 .... \otimes \sigma_1 \otimes \sigma_3 \otimes \mathbf{1} \\
\gamma_6 &= \sigma_1 \otimes \sigma_1 \otimes \sigma_1 .... \otimes \sigma_2 \otimes \mathbf{1} \otimes \mathbf{1} \\
\gamma_7 &= \sigma_1 \otimes \sigma_1 \otimes \sigma_1 .... \otimes \sigma_3 \otimes \mathbf{1} \otimes \mathbf{1}
\end{aligned}
\tag{A.4}
$$

$$..$$

$$
\begin{aligned}
\gamma_{d-3} &= \sigma_1 \otimes \sigma_1 \otimes \sigma_3 .... \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \\
\gamma_{d-2} &= \sigma_1 \otimes \sigma_2 \otimes \mathbf{1} .... \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \\
\gamma_{d-1} &= \sigma_1 \otimes \sigma_3 \otimes \mathbf{1} .... \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \\
\gamma_{d} &= \sigma_2 \otimes \mathbf{1} \otimes \mathbf{1} .... \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} \\
\gamma_{d+1} &= \sigma_3 \otimes \mathbf{1} \otimes \mathbf{1} .... \otimes \mathbf{1} \otimes \mathbf{1} \otimes \mathbf{1} = (-i)^{d/2} \gamma_1 \gamma_2 ... \gamma_d
\end{aligned}
$$

Each of the first d rows has $d/2$ two by two factors. Two inequivalent representations of the $2^{d/2-1}$ Euclidean Weyl matrices can be obtained by replacing in γ_μ the first "$\sigma_1 \otimes$" by 1 for $\mu = 1, .., d-1$ and "$\sigma_2 \otimes$" by $\mp i$ for $\mu = d$.

The overlap O, is given by

$$
O = <v_+|v_- >, \qquad \mathcal{H}^\pm |v_\pm > = E^\pm_{\min} |v_\pm > .
\tag{A.5}
$$

O is the regularized expression for $\det D(\xi)$. $E^\pm_{\min}$ denote minimal energies which define the associated states, assuming no degeneracies. The state $<v_+|$ becomes trivial with the choice $H^+ = H(-\infty)$. Choosing exactly one vector in $\mathcal{V}$ from the set $\{e^{i\Phi}|v_- > , 0 < \Phi \leq 2\pi\}$ for each gauge background, amounts to a "phase choice" for O.

Appendix B.

The main purpose of this appendix is to give a detailed derivation of (2.14).

On the lattice α replaces (μ, x, ij) where, for definiteness take $G_s = SU(n)$ with $i, j = 1, ...n$. ξ labels a point in the space $\prod_{\mu,x} G_s$. One value of ξ contains a complete description of the gauge background, namely $n^2 - 1$ real numbers per link.

Let us denote the fermionic operator $\partial_\alpha \mathcal{H}(\xi)$ by $R_\alpha(\xi)$. It is defined by:

$$\mathcal{H}(\xi + \delta\xi) - \mathcal{H}(\xi) = R_\alpha(\xi)\delta\xi_\alpha, \qquad (B.1)$$

where $\delta\xi_\alpha$ is an infinitesimal change in the gauge background, equivalent to $\delta U_\mu^{ij}(x)$. Define the quantities $(\delta U^g)_\mu^{ij}(x)$ for a finite lattice gauge transformation $g^{ij}(x)$:

$$(\delta U^g)_\mu^{ij}(x) = g^{\dagger\ ik}(x + \mu)\delta U_\mu^{kl}(x)g^{lj}(x). \qquad (B.2)$$

The $\delta U_\mu^{kl}(x)$ are restricted to a $n^2 - 1$ linear linear space per link by the linear requirement

$$U_\mu^\dagger \delta U_\mu + \delta U_\mu^\dagger U_\mu = 0. \qquad (B.3)$$

(B.2) ensures $U_\mu^{g\ \dagger}\delta U_\mu^g + \delta U_\mu^{g\ \dagger}U_\mu^g = 0$ if (B.3) holds. Equation (B.2) can be rewritten as:

$$(\delta\xi^g)_\beta = (\mathcal{D}^{-1}(g))_{\alpha\beta}\delta\xi_\alpha. \qquad (B.4)$$

This is the lattice equivalent to the variation of (2.1). Now,

$$\mathcal{H}((\xi + \delta\xi)^g) - \mathcal{H}(\xi^g) = \mathcal{H}(\xi^g + \delta\xi^g) - \mathcal{H}(\xi^g) = R_\alpha(\xi^g)(\delta\xi^g)_\alpha = R_\alpha(\xi^g)(\mathcal{D}^{-1}(g))_{\beta\alpha}\delta\xi_\beta. \qquad (B.5)$$

On the other hand,

$$\mathcal{H}((\xi + \delta\xi)^g) - \mathcal{H}(\xi^g) = G^\dagger(g)[\mathcal{H}(\xi^g + \delta\xi^g) - \mathcal{H}(\xi^g)]G^g = G^\dagger(g)R_\beta(\xi)G(g)\delta\xi_\beta. \qquad (B.6)$$

Comparing (B.5) and (B.6) gives, on account of the arbitrariness of $\delta\xi_\alpha$,

$$R_\alpha(\xi^g) = (\mathcal{D}(g))_{\alpha\beta}G^\dagger(g)R_\beta(\xi)G(g). \qquad (B.7)$$

This is equation (2.14).

References.

[1] L. Alvarez-Gaume, P. Ginsparg, Nucl. Phys. B 243 (1984) 449.

[2] R. Narayanan, H. Neuberger, Phys. Lett. B 302 (1993) 62.

[3] R. Narayanan, H. Neuberger, Nucl. Phys. B 443 (1995) 305.

[4] S. Randjbar-Daemi and J. Strathdee, Phys. Lett. B348 (1995) 543; Nucl. Phys. B443 (1995) 386; Nucl. Phys. B461 (1996) 305; Nucl. Phys. B466 (1996) 335.

[5] S. Randjbar-Daemi and J. Strathdee, Phys. Lett. B402 (1997) 134.

[6] A. Phillips, Annals of Physics 161 (1985) 399.

[7] M. V. Berry, Proc. R. Lond. A392 (1984) 45.

[8] W. Bardeen, B. Zumino, Nucl. Phys. B244 (1984) 421.

[9] B. Simon, Phys. Rev. Lett. 51 (1983) 2167.

[10] R. D. Ball, Phys. Rep. 182 (1989) 1.

[11] R. Narayanan, H. Neuberger, Nucl. Phys. B412 (1994) 574.

[12] K. Fujikawa, Phys. Rev. Lett. 42 (1979) 1195.

[13] H. Neuberger, Phys. Lett. B417 (1998) 141; hep-lat/9801031; Phys. Rev. D57 (1998) 5417.

[14] D. B. Kaplan, Phys. Lett. B288 (1992) 342.

[15] S. G. Naculich, Nucl. Phys. B296 (1988) 837.

[16] H. Sonoda, Nucl. Phys. B266 (1986) 410; A. J. Niemi, G. W. Semenoff, Phys. Rev. Lett. 55 (1985) 927.

[17] J. E. Avron, R. Seiler, B. Simon, Phys. Rev. Lett. 51, (1983) 51.

[18] R. Narayanan, H. Neuberger, Phys. Lett. B 393 (1997) 360; Y. Kikukawa, R. Narayanan, H. Neuberger, Phys. Lett. B 399 (1997) 105.

[19] Y. Kikukawa, R. Narayanan, H. Neuberger, Phys. Rev. D57 (1998) 1233.

[20] D. Foerster, H. B. Nielsen, M. Ninomiya, Phys. Lett. B 94 (1980) 135.

[21] R. Narayanan, H. Neuberger, Phys. Rev. Lett. 71 (1993) 3251.

[22] H. B. Nielsen, M. Ninomiya, NBI-HE-91-08, (1991), published in "Trieste conference on topological methods in quantum field theories", pages 149-171, (World Scientific, 1991, editor: W. Nahm).

[23] S. Pancharatnam, The Proceedings of the Indian Academy of Sciences, Vol. XLIV, No. 5, Sec. A, (1956); M. V. Berry, Journal of Modern Optics, 34, (1987) 1401.

[24] P. Hasenfratz, V. Laliena, F. Niedermayer, hep-lat/9801021.

[25] D. Boyanovsky, E. Dagotto, E. Fradkin, Nucl. Phys. B 285 (1987) 340; Y. Shamir, Nucl. Phys. B406 (1993) 90.